Course: ***Algorithms in the Real World (CS294-3)***

Instructor: ***Guy Blelloch (visiting from CMU)***
Soda 615

Times: ***Wednesdays and Fridays, 12:30-2:00***

URL: ***http://http.cs.berkeley.edu/~guyb/algs.html***

Yes, this course does cover the "theory" prelim breadth requirement.

Topics (approximate dates)

- Compression (Aug 29, Sept. 3, 5)
- Cryptography (Sept. 10, 12, 17)
- Linear Programming (Sept. 19, 24, 26)
- Integer Programming (Oct 1, 3)
- Triangulation and Meshing (Oct 8, 10, 15)
- Nbody Simulation (Oct 17, 22)
- VLSI Layout (Oct 24, 29, 31)
- Pattern Matching (Nov 5, 7)
- Indexing and Search Engines (Nov 12, 14, 19)
- Clustering (Nov 21, 26)

Requirements and Grades

Requirements

- Readings (many from the web, I'll hand out the rest)
- Scribe one class
- Assignments
- Take-home open-book final

Grades

- 50% Assignments
- 30% Final
- 20% Qualilty of scribe notes

Expected background

- Asymptotic analysis (solving recurrences, NP-complete problems)
- Graph terminology and algorithms (e.g. min-spanning-tree and max-flow algorithms)
- Dynamic programming
- Balanced trees (e.g. splay trees or 2-3 trees)
- Fast fourier transform

What the class will cover

Application areas that use sophisticated algorithms

Many algorithms

The transition from theory to practice

Various considerations and tradeoffs in algorithm design and deployment

- time vs. quality
- generality
- characterizing the input
- trust and momentum

Benchmark Suites

Why care about algorithms?

Just about every field uses sophisticated algorithms.

- engineering
- physical sciences
- social sciences
- business

The level of sophistication used in practice has greatly increased in recent years
Asymptotics becoming more important.

- Linear programming (interior point methods)
- Factoring (number field sieve)
- ....

Standard complaints about algorithm analysis

- Irrelevant problem
- Impractical constants hidden in big-O
- Bad machine models
- Too complicated

On Strassen

From Numerical Recipes (Press et al. 1986):

"We close this chapter with a little entertainment, a bit of algorithmic prestidigitation ....."

and later in the same section

"This is fun, but let's look at practicabilities: If you estimate how large N has to be before the difference between exponent 3 and exponent 2.807 is substantial enough to outweigh the bookkeeping overhead, arising from the complicated nature of the recursive Strassen algorithm, you will find that LU decomposition is in no immediate danger of becoming obsolete."

From an experimental paper (Bailey, Lee and Simon):

"For many years it was though that [the level at which Strassen' s algorithm is more efficient] was well over 1000x1000. In fact, for some new workstations, such as the Sun-4 and SGI IRIS 4D, Strassen is faster for matrices as small as 16x16."

Most modern matrix libraries now use Strassen, typically as a choice:

- Cray libraries
- IBM ESSL library
- LAPACK

Topics

Each topic will cover a problem area, along with associated algorithms and applications.

- Problem: Input/Output specification
- Algorithm: Method of solving a problem
- Application: A use of a problem
- Problem area: A set of related problems

Compression

Problems

- Lossless
- Lossy

Algorithms

- (Lossless) Huffman coding, LZW , LZ77, PPM, DMC
- (Lossy) T ransform coding, fractal coding, vector quant..

Applications

- communication (reducing bandwidth) --
- archiving
- encryption and machine learning

Cryptography

Problems

- Symmetric encryption (secret keys)
- Asymmetric encryption (public keys)
- Other: Digital signatures, hash functions

Algorithms

- (Symmetric) DES, NewDES, Skipjack, IDEA
- (Asymmetric) Knapsack, RSA, Diffie-Hellman, ElGamal
- (Hash functions) MD5, SHA, T iger

Applications

- Kerberos, Netscape, PGP , ISDN, Digital cash, ...

Linear Programming

Problems (and related problems)

- Straight linear programs
- Network flows
- Multiobjective optimization

Algorithms

- Simplex (many variants)
- Ellipsoid and interior point algorithms

Applications

- Finance, job-shop scheduling, inventory control
- A step in integer programming

Integer programming

Problems

- General
- 0-1 programming

Algorithms

- Approximating with linear programming
- Branch-and-bound, Cutting plane

Applications

- Airline scheduling (Delta, American, United)
- Delivery and inventory management (W allmart, Shell)
- VLSI placement

Triangulation

Problems

- Delaunay and constrained-Delaunay
- Boundary meshing (conforming Delaunay)

Algorithms

- Guibas-Stolfi, Dwyer , Bucketing, Edge-flipping, Edelsbruner
- Bern-Epstein-Gilbert, Rupert (boundary meshing)

Applications

- Meshing for finite-element simulations (airflow , stress, ...)
- Terrain modeling
- Interpolation and compression

VLSI Layout

Problems

- Placement
- Global and detailed routing

Algorithms

- (Placement) partition-based, cluster -growth, simulated-annealing
- (Routing) Maze, Shortest path, Steiner tree, integer prog.

Applications

- Intel, Sun, SGI, HP , ...
- FPGAs

N-body problem

Problems

- Coulomb forces
- Other forces

Algorithms

- particle-mesh, nested-grid particle-mesh
- (tree based) Barnes-Hut, Greengard, Callahan

Applications

- Astronomy -- formation of galaxies
- Biology/chemistry -- molecular dynamics
- Other: plasma physics, PDEs, Fluid-dynamics

Pattern matching

Problems

- Exact matching
- Approximate matching

Algorithms

- (Exact) Knuth-Morris-Pratt, Boyer -More, Rabin-Karp
- (Approximate) Smith-W aterman

Applications

- DNA sequence analysis
- Text analysis

Search Engines

Problems

- Exact vs. Partial matching
- Image and sound retrieval
- Grouping

Algorithms

- Inverted indices
- Clustering, Filtering

Applications

- Hotbot, L ycos, Altavista, ...
- Searching in multimedia digital libraries

Clustering

Problems

- Fixed cardinality vs. not
- Predefined groupings vs. not

Algorithms

- Hierarchical (top down or bottom up)
- Optimization, density search, clumping, SVD

Applications

- Compression, search-engines
- Classification (e.g. V isa)
- Machine learning